

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-268952
 (43)Date of publication of application : 20.09.2002

(51)Int.Cl.

G06F 12/16
 G06F 3/06
 G06F 12/00
 G06F 12/14
 G06F 15/16
 G06F 15/17

(21)Application number : 2001-070897

(71)Applicant : MITSUBISHI HEAVY IND LTD

(22)Date of filing : 13.03.2001

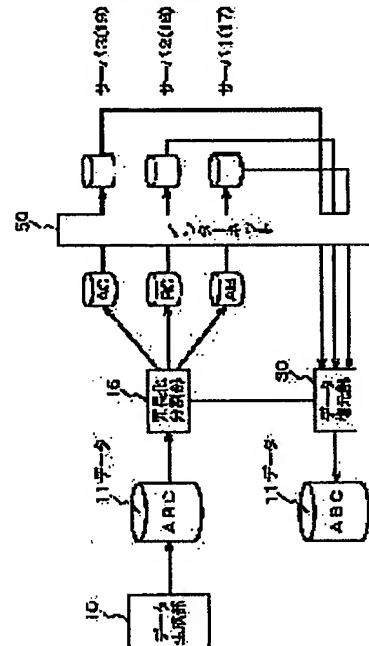
(72)Inventor : EMOTO HIDEAKI
 SAGAWA ISAO

(54) DISTRIBUTED STORAGE SYSTEM OF DATA, PROGRAM AND SERVER

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a distributed storage system of data by which the data can be restored from the data stored in a server other than a certain server which is break down even when the certain server is break down.

SOLUTION: The distributed storage system of data is provided with a data dividing part 15 to divide a single piece of data 11 into a plurality of pieces of divided data, N (N is plural) servers 17, 18, 19 to store the plurality of pieces of the divided data respectively and a data restoring part 30 to restore the single piece of the data based on the plurality of pieces of the data stored in the (N-1) servers 17, 18.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-268952

(P2002-268952A)

(43)公開日 平成14年9月20日 (2002.9.20)

(51) Int.Cl. ⁷	識別記号	F I	テ-マコ-ト ⁸ (参考)
G 06 F 12/16	3 2 0	G 06 F 12/16	3 2 0 L 5 B 0 1 7
	3 1 0		3 1 0 M 5 B 0 1 8
3/06	5 4 0	3/06	5 4 0 5 B 0 4 5
12/00	5 1 4	12/00	5 1 4 E 5 B 0 6 5
	5 3 1		5 3 1 D 5 B 0 8 2

審査請求 未請求 請求項の数10 O L (全7頁) 最終頁に続く

(21)出願番号 特願2001-70897(P2001-70897)

(22)出願日 平成13年3月13日 (2001.3.13)

(71)出願人 000006208

三菱重工業株式会社

東京都千代田区丸の内二丁目5番1号

(72)発明者 江本 英晃

兵庫県高砂市荒井町新浜2丁目1番1号

三菱重工業株式会社高砂製作所内

(72)発明者 佐川 功

兵庫県高砂市荒井町新浜2丁目1番1号

三菱重工業株式会社高砂製作所内

(74)代理人 100102864

弁理士 工藤 実 (外1名)

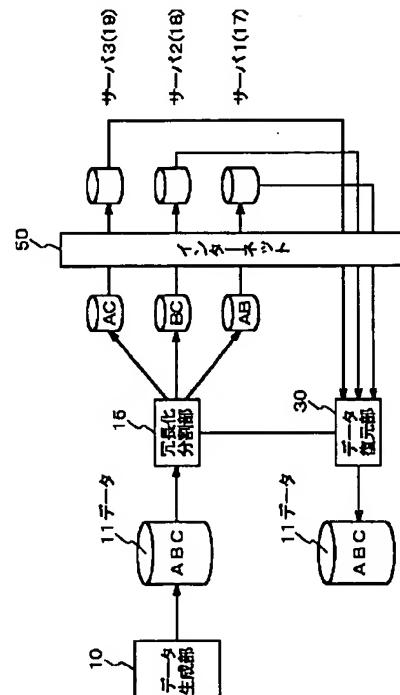
最終頁に続く

(54)【発明の名称】 データの分散保存システム、プログラムおよびサーバ

(57)【要約】

【課題】 或るサーバがダウンした場合であってもそれ以外のサーバに保存されたデータからデータが復元することができるデータの分散保存システムを提供する。

【解決手段】 単一のデータ11を複数の分割データに分割するデータ分割部15と、前記複数の分割データをそれぞれ格納するN個 (Nは複数) のサーバ17、18、19と、前記 (N-1) 個の前記サーバ17、18に格納された前記複数の分割データに基づいて、前記单一のデータを復元するデータ復元部30とを備えている。



【特許請求の範囲】

【請求項1】 単一のデータを複数の分割データに分割するデータ分割部と、前記複数の分割データをそれぞれ格納するN個（Nは複数）のサーバと、前記（N-1）個の前記サーバに格納された前記複数の分割データに基づいて、前記单一のデータを復元するデータ復元部とを備えたデータの分散保存システム。

【請求項2】 請求項1記載のデータの分散保存システムにおいて、前記サーバは、インターネット・ディスクスペース供給機器であるデータの分散保存システム。

【請求項3】 請求項1または2に記載のデータの分散保存システムにおいて、前記データ分割部は、前記複数の分割データが冗長性を有するように、前記单一のデータを分割するデータの分散保存システム。

【請求項4】 請求項1または2に記載のデータの分散保存システムにおいて、前記データ分割部は、RAID（Redundant Arrays of Inexpensive Disks）方式により、前記单一のデータを分割するデータの分散保存システム。

【請求項5】 単一のデータを入力するステップと、前記入力された单一のデータを複数の分割データに分割するステップと、前記複数の分割データをそれぞれに複数のインターネット・ディスクスペース供給機器に出力するステップの各ステップをコンピュータに実行させるためのプログラム。

【請求項6】 複数のインターネット・ディスクスペース供給機器にそれぞれ格納された、单一のデータが分割されてなる複数の分割データを前記複数のインターネット・ディスクスペース供給機器から入力するステップと、前記入力した複数の分割データに基づいて、前記单一のデータを復元するステップの各ステップをコンピュータに実行させるためのプログラム。

【請求項7】 単一のデータを入力するステップと、前記入力された单一のデータを複数の分割データに分割するステップと、前記複数の分割データをそれぞれにN個（Nは複数）のサーバに出力するステップと、前記（N-1）個の前記サーバに格納された前記複数の分割データを前記（N-1）個のサーバから入力するステップと、前記入力した複数の分割データに基づいて、前記单一のデータを復元するステップの各ステップをコンピュータに実行させるためのプログラム。

【請求項8】 請求項7記載のプログラムにおいて、

10 【請求項10】 請求項9記載のサーバにおいて、前記サーバは、インターネット・ディスクスペース供給機器であるサーバ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データの分散保存システム、プログラムおよびサーバに関し、特に、データの秘匿性および復元性に優れたデータの分散保存システム、プログラムおよびサーバに関する。

【0002】

20 【従来の技術】インターネットのサーバ等のデータ保管システムにデータを保存する場合、以下の2つの問題がある。

（1）サーバの管理者は、預かったデータを覗くことができるため、機密性に欠ける。

（2）サーバがメンテナンスやトラブルの場合、データハンドリングができなくなる。すなわち、サーバのデータ保管システムには、必ずメンテナンスの時期があり、そのときには、そのサーバに保存されているデータをアクセスして参照したり、データを入出力することができなくなる。また、サーバが震災などによりダウンした場合にも、同様にデータアクセスやデータ入出力が不可能になる。

【0003】

【発明が解決しようとする課題】従来は、上記（2）の問題を回避するために、複数のサーバに同じデータを分散保存して、或るサーバがダウンした（メンテナンスやトラブル）場合であっても、それ以外のサーバに保存されたデータのハンドリングができるようになっていた。この方法によれば、同じデータが保存された複数のサーバのうちいずれか一つのサーバが稼動中であれば、そのサーバとの間でデータハンドリングを行うことができる。その反面、いずれのサーバの管理者も、自社のデータ保管システムに保存されたデータからデータ内容を認識することができるため、上記（1）で述べた通り、機密性に欠けるという問題がある。

40 【0004】機密性が十分に確保されると共に、或るサーバがダウンした場合であってもそれ以外のサーバに保存されたデータからデータが復元することができるデータの分散保存システム、プログラムおよびサーバが望まれる。

【0005】なお、特開平11-238076号公報には、次の空間情報流通システムが開示されている。複数の空間情報を格納、提供する1以上の空間情報提供計算機と、前記空間情報提供計算機から提供される複数の空間情報を利用する1以上の空間情報利用計算機とが、ネットワークで接続されている空間情報流通システムにおいて、前記空間情報利用計算機における複数の空間情報の利用履歴である利用形態情報を有することを特徴とする。

【0006】また、特開2000-259359公報には、次のRAID装置が開示されている。データを複数のデータ用ディスクに分散保存するRAID装置において、前記複数のデータ用ディスクに複数台のパリティ用ディスクを冗長する冗長手段と、前記データ用ディスクのデータ保存領域およびパリティ用ディスクのパリティ保存領域をビット単位に任意整数にブロック分けするブロック分割手段と、拡大ガロア体GF(2ⁿ) (nは整数)を用いて、前記全ディスクの所定ブロックどうしで連なるグループ単位ごとに前記データ用ディスクに付加するパリティ用ディスクのパリティを算出するパリティ算出手段および前記任意のデータ用ディスク障害時、障害発生ディスクのデータを未知データとし、前記拡大ガロア体GF(2ⁿ)で定める規則に従って連立合同式を作成し、この連立合同式から前記未知データを復元するデータ復元手段を有するRAID制御装置とを備えたことを特徴としている。

【0007】本発明の目的は、機密性が十分に確保されると共に、或るサーバがダウンした場合であってもそれ以外のサーバに保存されたデータからデータが復元することができるデータの分散保存システム、プログラムおよびサーバを提供することにある。本発明の他の目的は、機密性が十分に確保されると共に、或るサーバがダウンした場合であってもそれ以外のサーバに保存されたデータからデータが復元することができるデータの分散保存システム、プログラムおよびサーバを提供することにある。

【0008】

【課題を解決するための手段】その課題を解決するための手段が、下記のように表現される。その表現中の請求項対応の技術的事項には、括弧()つき、番号、記号等が添記されている。その番号、記号等は、請求項対応の技術的事項と実施の複数・形態のうちの少なくとも一つの形態の技術的事項との一致・対応関係を明白にしていくが、その請求項対応の技術的事項が実施の形態の技術的事項に限定されることを示されるためのものではない。

【0009】本発明のデータの分散保存システムは、単一のデータ(11)を複数の分割データに分割するデータ分割部(15)と、前記複数の分割データをそれぞれ格納するN個(Nは複数)のサーバ(17、18、1

9)と、前記(N-1)個の前記サーバ(17、18)に格納された前記複数の分割データに基づいて、前記单一のデータ(11)を復元するデータ復元部(30)とを備えている。

【0010】本発明のデータの分散保存システムにおいて、前記サーバ(17、18、19)は、インターネット・ディスクスペース供給機器である。

【0011】本発明のデータの分散保存システムにおいて、前記データ分割部(15)は、前記複数の分割データが冗長性を有するように、前記单一のデータ(11)を分割する。

【0012】本発明の分散保存システムにおいて、前記データ分割部(25)は、RAID (Redundant Arrays of Inexpensive Disks) 方式により、前記单一のデータ(11)を分割する。

【0013】本発明のプログラムは、単一のデータ(11)を入力するステップと、前記入力された単一のデータ(11)を複数の分割データに分割するステップと、前記複数の分割データをそれぞれに複数のインターネット・ディスクスペース供給機器(17、18、19)に出力するステップの各ステップをコンピュータに実行させるためのプログラムである。

【0014】本発明のプログラムは、複数のインターネット・ディスクスペース供給機器(17、18、19)にそれぞれ格納された、単一のデータ(11)が分割されてなる複数の分割データを前記複数のインターネット・ディスクスペース供給機器(17、18、19)から入力するステップと、前記入力した複数の分割データに基づいて、前記单一のデータ(11)を復元するステップの各ステップをコンピュータに実行させるためのプログラムである。

【0015】本発明のプログラムは、単一のデータ(11)を入力するステップと、前記入力された単一のデータ(11)を複数の分割データに分割するステップと、前記複数の分割データをそれぞれにN個(Nは複数)のサーバ(17、18、19)に出力するステップと、前記(N-1)個の前記サーバ(17、18、19)に格納された前記複数の分割データを前記(N-1)個のサーバ(17、18、19)から入力するステップと、前記入力した複数の分割データに基づいて、前記单一のデータ(11)を復元するステップの各ステップをコンピュータに実行させるためのプログラムである。

【0016】本発明のプログラムにおいて、前記サーバは、インターネット・ディスクスペース供給機器である。

【0017】本発明のサーバは、他の複数の外部サーバ(18、19)とともに単一のデータ(11)を分割保存するサーバ(17)であって、前記サーバ(17)は、前記单一のデータ(11)が第1、第2および第3

分割データに分割されたとき、前記第1、第2および第3分割データのうち前記他の複数の外部サーバ(18、19)がそれぞれに格納する前記第1および第2分割データ以外の前記第3分割データを格納する。

【0018】本発明のサーバにおいて、前記サーバ(17)は、インターネット・ディスクスペース供給機器である。

【0019】

【発明の実施の形態】以下、添付図面を参照して、本発明のデータ分散保存システムの一実施形態について説明する。図1は、本実施形態のデータ分散保存システムの構成を示すブロック図である。

【0020】図1に示されるように、符号10は、データ生成部である。データ生成部10は、単一のデータ11を生成する。そのデータ11は、冗長化分割部15に送られる。

【0021】冗長化分割部15は、データ11を冗長性を持たせて分割し、それらの分割したデータを複数のデータ保管システム17、18、19にそれぞれ割り当てて保存させる。

【0022】例えば、データ生成部10がデータ11として「A」、「B」、「C」の3つのブロックからなる単一のデータ「ABC」を生成した場合、冗長化分割部15は、そのデータ「ABC」を冗長性を持たせて「AB」、「BC」、「AC」の3つに分割する。そして、冗長化分割部15は、その分割したデータ「AB」を第1データ保管システム(サーバ)17に割り当てて保存させ、同様に、データ「BC」を第2データ保管システム(サーバ)18に割り当てて保存させ、データ「AC」を第3データ保管システム(サーバ)19に割り当てて保存させる。

【0023】この場合、冗長化分割部15は、データ11を冗長性を持たせて分割した後、その分割した各データを、インターネット50上でFTP(File Transfer Protocol)、HTTP(Hypertext Transfer Protocol)という転送プロトコルを用いて、それぞれのデータ保管システム(サーバ)17、18、19に転送させることができる。

【0024】各データ保管システム(サーバ)17、18、19は、「ABC」からなるデータ11の全体を保存しているわけではなく、いずれもその一部が欠落したデータを保存しているに過ぎないため、データ内容を認識することができず、これにより、データの機密性が確保される。

【0025】データ復元部30は、上記3つのデータ保管システム(サーバ)17、18、19のうち1つがメンテナンスやトラブルなどでダウンした場合にも、残りの2つのデータ保管システム(サーバ)に保存されたデータから、「ABC」からなるデータ11の全体を復元

することができる。

【0026】この場合、データ復元部30は、インターネット50上でFTP、HTTPという転送プロトコルを用いて、それらのデータ保管システム(サーバ)17、18、19からデータ復元部30まで各データを転送させることができる。

【0027】データ復元部30は、上記のデータ復元を行う前提として、冗長化分割部15からデータ分割手法および/または分割されたデータの保存先(どのデータ保管システム(サーバ)のどのメモリエリアか)を示すデータが通知されている。データ復元部30は、そのデータに基づいて、保存先のデータ保管システム(サーバ)から所望のデータを読み出して復元する。

【0028】ここで、冗長化分割部15がデータを分割する数は3に限らず、保存させるべき複数のデータ保管システムの数に等しい2以上の数である。

【0029】次に、図2を参照して、本実施形態の変形例を説明する。図1と同じ構成要素については同じ符号を付してその詳細な説明を省略する。

【0030】本例では、冗長化分割部15に代えて、データ分割部25が設けられている。このデータ分割部25は、冗長化分割部15のようにデータ11を冗長性を持たせて分割するのではなく、後述するRAID5方式によりデータ11を分割する。すなわち、データ分割部25は、「ABC」からなる単一のデータ11を、「A」、「B」、「C」、「P(パリティ)」の4つに分割して、それぞれをサーバ1~4(17~20)に保存させる。

【0031】ここで、「P(パリティ)」は、「A」、「B」、「C」の3つのうちの2つ(例えば「A」、「C」)との間で排他的論理和をとることで残りの1つ(この例では「B」)を求めることができるものである。すなわち、データ復元部30は、「A」、「B」、「C」をそれぞれ保存するサーバ1~3(17~19)のいずれ1つがダウンしても、「P」を用いることで、そのダウンしたサーバに保存されたデータを求めることができ、これによりデータ11を復元することができる。

【0032】この場合、データ分割部25による分割方法は、RAID5には限定されず、後述するRAID5以外のRAIDの方法であることができる(但し、後述する理由でRAID1は単独では採用できない)。

【0033】一般に、RAID(Redundant Arrays of Inexpensive Disks)装置は、各種のデータを複数の安価なディスクに分散し保存することにより、高性能ディスクと同等の性能を得る装置である。しかし、小型のディスクを多数使用すれば、ディスクの故障が増加し、それだけデータが消失する危険性も高くなるので、単にディスクの数を増やすだけでなく、余分なディスクを用意し、冗長性をも

たせることにより、安定性および高性能化を図っている。

【0034】R A I D 装置は、データ用ディスクの他に、冗長性をもたせるためのパリティ用ディスクが設けられ、データ用ディスクに障害が発生したとき、冗長性をもったパリティデータを用いて、障害によって失われたデータを復元することが行われている。

【0035】R A I D 装置は、ディスクの組合せないし負荷分散の観点から、大まかには6つのレベルR A I D 1～R A I D 6に分類され、ディスク障害時のデータの復元化を図っている。

【0036】R A I D 1は、ミラーリング (m i r r o r i n g) とも呼ばれる方式であって、この方式は、データ用ディスクが2つのグループに分けられ、同一のデータを2つのグループデータディスクに保存する二重化データ保存方式である。データの書き込みは両グループ同時に、データの読み出しあはれかのデータ用ディスクから行う。その結果、一方のグループデータ用ディスクの障害時、別のグループに切替えるだけで消失データを簡単に復元できる。

【0037】上記のR A I D 1は、同一のデータが複数のデータ保管システムに保存されるものであり、各データ保管システムが完全な形の（データの欠落が無い）データ11（上記例では「A B C」）を保存するため、データの機密性に欠ける。よって、本実施形態およびその変形例では、単独では採用することができない。

【0038】但し、R A I D 1を他の分割方法と組み合わせて用いることは可能である。すなわち、図1に示した例でいえば、分割されたデータ「A B」を第1データ保管システム17に割り当てて保存させるとともに、データ「A B」をミラーリングさせて、第4のデータ保管システム（図示せず）にも同じくデータ「A B」を保存させることができる。

【0039】R A I D 2は、データを複数のディスクにビット単位で分散記録する一方、コンピュータのメモリで利用されているエラー訂正コード（E C C）を付加し、データ用ディスクの障害により失われたデータについて、エラー訂正コードを用いて復元し信頼性を高める方式である。

【0040】R A I D 3は、ディスクごとに障害を検知する仕組みがあることを前提とし、E C Cを使わずにデータをビット単位で複数のディスクに分散させ、パリティ用ディスク1台を追加し、データ用ディスクの障害により失われたデータについて、ビット単位の排他的論理和によって障害データ用ディスク1台を復元する方式である。

【0041】R A I D 4は、ディスクごとに障害を検知する仕組みがあることを前提とし、R A I D 3がデータをビット単位で分散させたのに対し、データをブロック単位で複数のデータ用ディスクに分散させ、パリティ用

ディスク1台を追加し、障害により失われたデータについて、ブロック単位の排他的論理和によって障害データ用ディスク1台を復元する方式である。

【0042】R A I D 5は、同じくディスクごとに障害を検知する仕組みがあることを前提とし、パリティをブロック単位に全てのディスクに分散保持させ、障害により失われたデータは、ブロック単位の排他的論理和によって障害データ用ディスク1台を復元する方式である。

【0043】R A I D 6は、R A I D 5が障害対策に单一パリティを用いたのに対し、パリティを2次元に拡張したり、R e e d - S o l o m o n コードを用いてエラー訂正機能を強化することにより、複数台のデータ用ディスク障害を復元し、信頼性を高める方式である。

【0044】また、本実施形態およびその変形例において、複数のデータ保管システムは、同一法人であると、それらの複数のデータ保管システム相互間での連携が可能であり、それらの複数のデータ保管システムのそれぞれに分散保存された各データを組み合わせてデータ内容を認識してしまうおそれがある。そのため、互いに関連性の無い別法人のデータ保管システムに分散保存させることが望ましい。

【0045】さらに、本実施形態およびその変形例において、複数のデータ保管システムは、同一地域に集中して存在していると、万一、天変地異（震災など）が発生したときにその地域の複数のデータ保管システムがまとめてダウンするおそれがあるため、互いに異なる地域に設置されることが望ましい。

【0046】なお、本実施形態およびその変形例では、データ保管システムがインターネットのサーバであるとして説明したが、本発明ではインターネットのサーバに限定されるものではない。複数のデータ保管システムの一つが、サーバではなく自社のサーバであり、他のデータ保管システムがサーバであることができる。また、複数のデータ保管システムの一つが、会社の或る事業場内のサーバであり、他のデータ保管システムがその会社の国内外の営業所、関連・協力会社のサーバであることができる。

【0047】ここで、本実施形態と、従来一般のディスク・アレイ（R A I D）装置との相違について説明する。

【0048】従来一般のディスク・アレイ（R A I D）装置は、単独で（自社内などの閉じた領域に）存在するデータ保存システムにおいて実施され、その目的は専らデータ用ディスク障害時のデータ復元であった。すなわち、従来一般のディスク・アレイ（R A I D）装置においては、複数のディスクは、いずれも自社の管理下に設置されるため、機密性の確保という要請は殆ど無い。

【0049】これに対し、本実施形態およびその変形例では、インターネット50というパブリックでオープンなネットワーク上のサーバ（データ保存システム）にデ

ータを保存することを一前提としているため、従来では問題視され得なかった機密性確保の問題が初めて生じたのである。この機密性の問題を解決するために、データ11の全容が認識されないように、データ11の一部が欠落する形にデータ分割を行い、それぞれの分割データを各サーバに分散保存させているのである。また、そもそも本実施形態およびその変形例において、データ分割の方法はディスク・アレイ(RAID)方式に限定されるものではない(図1参照)。

【0050】また、従来一般のディスク・アレイ(RAID)装置は、あくまで自社の管理下に設置された「ディスクの故障」という状況が想定されたものに過ぎない。これに対し、本実施形態が解決しようとしている、サーバがメンテナンス時期などでデータハンドリングが一時的に不可能な状況、または或る地域に設置されたサーバだけが震災等により破壊された状況は、自社の管理下の「ディスクの故障」という状況とは次元が異なる。本実施形態が解決しようとしている上記状況に備えて、RAIDの分散保存の考え方の一部を適用したこと自体が従来無い新しいものである。

【0051】さらに、従来一般のディスク・アレイ(RAID)装置では、複数のディスクの一つが故障したときにもデータを復元することができるものの、それらのディスクからデータを読み出す装置または複数のディスクを制御するコントローラが故障すると、もはやディスクにアクセスできずにデータハンドリングが不可能になる。

【0052】これに対し、本実施形態によれば、インターネット50上の複数のサーバにアクセスできるのは特定の読み出し装置またはコントローラ(データ復元部30)に限定されないため、或る読み出し装置またはコントローラ(データ復元部30)が故障したときでも、イン*

* ターネット50に接続された他の読み出し装置またはコントローラ(図示しないデータ復元部)を用いることで、データハンドリングが可能である。

【0053】本実施形態において、データ保存システムの保存領域がプロバイダ(サーバ)から借りたディスクスペースであるとした場合には、通常、プロバイダからは24時間体制のディスクメンテナンスサービスが得られることから、データ保存システムが自社のサーバである場合に比べて、人件費をはじめとするメンテナンスにかかる諸経費の削減効果が得られる。

【0054】

【発明の効果】本発明のデータの分散保存システムによれば、機密性が十分に確保されると共に、或るサーバがダウンした場合であっても、それ以外のサーバに保存されたデータに基づいて、データを復元することができる。

【図面の簡単な説明】

【図1】図1は、本発明のデータの分散保存システムの一実施形態を示すブロック図である。

20 20 【図2】図2は、本発明のデータの分散保存システムの一実施形態の変形例を示すブロック図である。

【符号の説明】

10 データ生成部

11 データ

15 冗長化分割部

17 サーバ(データ保管システム)

18 サーバ(データ保管システム)

19 サーバ(データ保管システム)

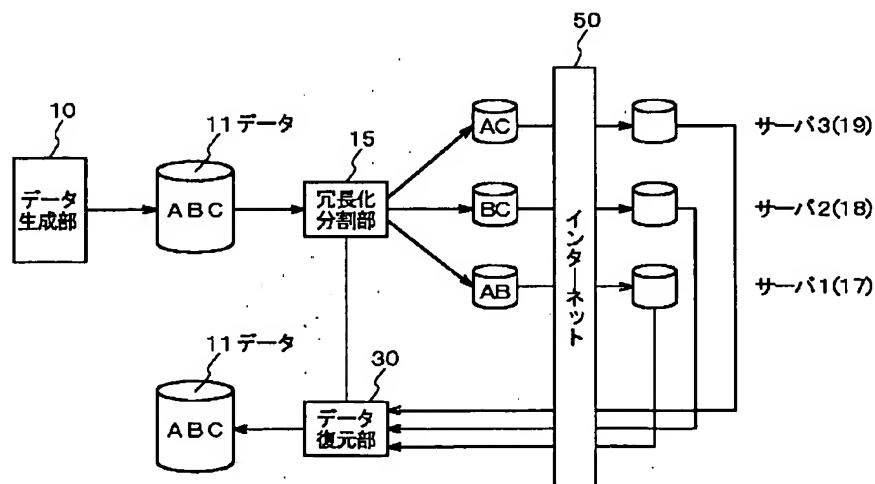
20 サーバ(データ保管システム)

25 データ分割部

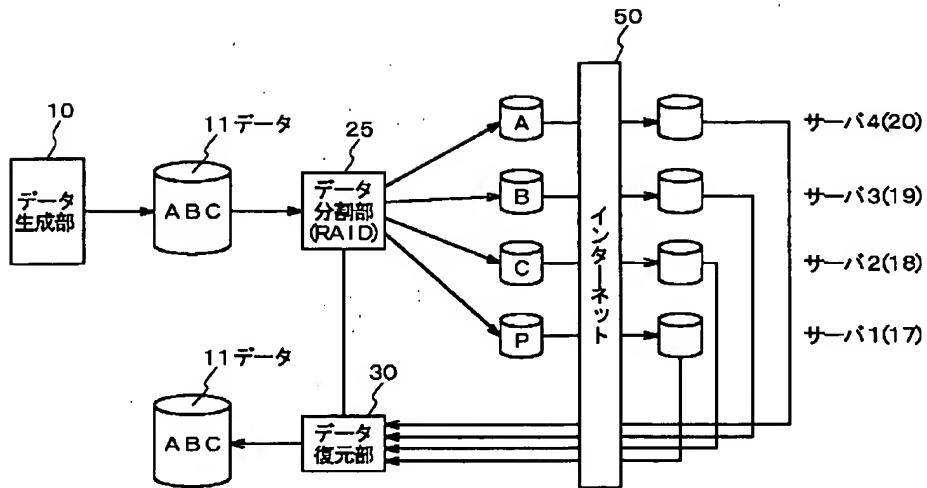
30 データ復元部

50 インターネット

【図1】



【図2】



フロントページの続き

(51) Int.C1.		識別記号	F I	マーク (参考)
G 0 6 F	12/00	5 4 5	G 0 6 F	12/00
	12/14	3 2 0		12/14
	15/16	6 2 0		15/16
	15/177	6 7 8		15/177
				5 4 5 A
				3 2 0 A
				6 2 0 H
				6 7 8 C

Fターム(参考) 5B017 AA03 BA10 CA07
 5B018 GA04 HA05 HA35 KA03 KA21
 MA12
 5B045 DD16 JJ43
 5B065 BA01 CA30
 5B082 DE06